Classification of Individuals at Risk for Alcoholism using Non-matching ERPs based on Wavelet Statistic Features and Artificial Neural Network

SH. Shobeihi[1], A.M. Nasrabadi[2], M.H. Moradi[3]
[1]Islamic AZAD University, Science & Research Branch, Tehran, Iran
[2]Biomedical Engineering Dept., Engineering Faculty, Shahed University, Tehran, Iran
[3] Biomedical Engineering Faculty, Amirkabir University of Technology, Tehran, Iran
e-mail: sh_shobeihi@yahoo.com

Abstract-**In this work we using Artificial Neural Network as classifier for classification of the group of individuals at high risk for alcoholism (HR) from ones at low risk for alcoholism (LR) based on Non-matching ERPs signals of Ingber database. By choosing the statistics features from wavelet domain and reduce the dimensionality of the features vector and using Multi Layer Perceptron network by Permute Cross Validation training, we can correctly classify those two groups over 88%.**

Keywords- **ERP, Alcoholism, Wavelet, Artificial Neural Networks, Multi Layer Perceptron, Cross Validation.**

I. Introduction

It is 60 years since Berger demonstrated electroencephalographic abnormalities in a histologically confirmed case of Alzheimer's disease. Since then, the electroencephalograph (EEG) has been widely used in the investigation of patients with suspected organic brain disease [1].

Alcoholism is a disease that runs in families and results at least in part from genetic risk factors. Children of alcoholics (COA's) are at higher risk than the general population for developing alcoholism. Evidence suggests that this risk is influenced by genetic factors. Researchers have identified several biological traits that appear to be genetically transmitted along with vulnerability to alcoholism. These traits can serve as markers to identify persons at risk and can provide valuable information on the development of alcohol use disorders. Because the processes of addiction occur largely in the brain, many studies have investigated various measures of brain function and much of this research has focused on the brain's electrical activity [2], [3], [4].

It is widely recognized that alcoholics manifest brain damage/dysfunction, and electrophysiological methods have long been used to elucidate the nature of this brain dysfunction.These neuroelectric phenomena may be recorded with the continuous electroencephalogram when the subject is at rest and not involved in a task or with the time-specific event-related potentials (ERPs) during cognitive tasks [5].

Event-related or evoked potentials (ERPs or EPs) consist of transient voltage changes that occur in response to a sensory stimulus. These take the form of a series of negative and positive waves [1]. In the ERP studies of alcoholism, reduced voltage of an ERP called P300, or P3, appears to characterize offspring of alcoholic families, regardless of whether the offspring are themselves alcoholic. Reduced P3 may indicate susceptibility to alcoholism and may elucidate mechanisms of alcohol's effects on the nervous system [3]; Also studies shown that HR (High risk for alcoholism) subjects produced smaller P300 amplitudes than LR (low risk for alcoholism) subjects for the visual matching tasks [6].

A time-frequency analysis of P300 which aimed to further understand the nature of the event-related oscillation (ERO) components which form the P300 wave and how these components may be used to differentiate alcoholic individuals from controls, demonstrate that in a similar way to the P300 amplitude, the condition of reduced theta and delta ERO power may precede the development of alcoholism and therefore represent a trait marker for it [7].

Functional MRI were collected during the performance of a visual oddball task, from LR subjects with high P300s and HR subjects with low P300s. Analysis of the fMRI data revealed two areas with significantly lower activation in the HR group when compared to the LR group: the bilateral inferior parietal lobule that showed significantly lower activation in the HR group in contrast to the LR group, and inferior frontal gyrus that it was not activated in the HR group but was only activated in the LR group. This finding indicates that a dysfunctional frontoparietal circuit may underlie the low P300 responses seen in HR subjects. This perhaps implies a deficiency in the rehearsal component of the working memory system [8].

As saw in the above, by recording ERPs under conditional tasks and study them, can show the differences between ERPs components in HR and LR groups. Our hypothesis assessed that by knowing these differences and using artificial neural network and wavelet transform as computational and signal processing tools, it's possible to distinguishing between the HR and LR groups. Also we will show that by using features extracted from Non-matching study and also reducing in features dimensionality and improving in MLP training phase by increasing its learning performance, we can improve the results of the classifier accuracy and also can use more than one electrode's features as input of the MLP without loosing the accuracy for better classification of those two groups.

II. Methodology

A. ERP Data and Recording

The data used in this work arises from a large study to examine EEG correlates of genetic predisposition to alcoholism, and were available for Lester Ingber in [9]. These data were collected by Henri Begleiter and associates at the Neurodynamics Laboratory at the State University of New York Health Center at Brooklyn, and prepared by David Chorlian. The subjects in this study consisted of a group at high risk for developing alcoholism and a low risk group. In this experiment 122 subjects participated.

The subject was seated in reclining chair located in a sound-attenuated RF shield room and fixated a point in the center of a computer display located 1 m away from their eyes. EEG activity was recorded using a 61-channel electrode cap (ECI, Electrocap International), which include 19 electrodes of the 10-20 International System and 42 additional electrode sites (Electrode Position Nomenclature, American Electroencephalographic Association, 1991).

All scalp electrodes were referred to Cz. Subjects were grounded with a nose electrode. Two additional bipolar deviations were used to record EOG. The signals were amplified with a gain of 10,000 with a bandpass between 0.02 and 50 Hz, and recorder on a Concurrent 55/50 computer. The amplified signals were sampled at a rate of 256 Hz during an epoch of 190 ms of pre-stimulus baseline and 1440 ms following each stimulus presentation. Trials with excessive eye and body movements (>73.3 uV) were rejected on-line.

To elicit the ERP, a modified delayed matching-to-sample task was used in which two picture stimuli appeared in succession with a 1.6s fixed inter-stimulus interval. The duration for the first (S1) and the second (S2) picture stimulus in each test trial was 300 ms. The interval between each trial was fixed to 3.2 s. All pictures were paired into two conditions, that is, matching and non-matching. In the matching condition, S1 was repeated as S2. In the non-matching condition, the S1 was followed by a picture that was completely different from S1 in terms of its semantic category.

The subject's task was to decide whether the second picture (S2) was the same as the first stimulus (S1). They were asked to press a mouse key in one hand if the S2 matched S1 and to press a mouse key on the other hand if the S2 differed from S1 after the presentation of S2 on each trial. ERPs were averaged only on artifact-free trials with correct responses for two cases, match S2 and non-match S2.

In the experiment, the subject was stimulated with pictures that were chosen from 1980 Snodgrass and Vanderwat picture set to identified an ERP component correlate with visual memory.

This experiment yields an ERP waveform consisting of three components which are most clearly discernible at the posterior electrodes. Component 1 (c110) ranging between 100 and 125 ms, component 2 (c175) ranging between 160 and 190 ms, and component 3 (c247) ranging between 220 and 260 ms [10].

B. ERP Data extracted from database

We selected subjects at high risk for alcoholism (HR) consisted 40 individuals, and subjects at low risk for alcoholism (LR) consisted 40 individuals, randomly from the available database [9]. Also for our work we selected just the signals recorded for Non-matching studies.

C. Wavelet transform and DWT

The Wavelet Transform (WT) gives a time-frequency representation of a signal that has two main advantages methods: (a) an optimal resolution both in the time and in the frequency domains; and (b) the lack of the requirement of stationarity of the signal. It is defined as the convolution between the signal *x(t)* and the wavelet functions $\psi_{a,b}(t)$

$$W_{\psi} X(a,b) = \langle x(t) \mid \psi_{a,b}(t) \rangle \tag{1}$$

where $\psi_{a,b}(t)$ are dilated (contracted) and shifted versions of a unique wavelet function $\psi(t)$

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \tag{2}$$

(a, b are the scale and translation parameters, respectively). The WT gives a decomposition of *x(t)* in different scales, tending to be maximum at those scales and time locations where the wavelet best resembles *x(t)*.

Since the parameters (a, b) are continuous valued, the transform is called continuous wavelet transform. In general, the scale and shift parameters of the discreet wavelet family are given by

$$a = a_0^j, \quad b = kb_0 a_0^j \tag{3}$$

where j and k are integers. The function family with discretized parameters becomes

$$\psi_{j,k}(t) = a_0^{-j/2} \psi(a^{-j} t - kb_0) \tag{4}$$

$\psi_{j,k}(t)$ is called the discrete wavelet transform (DWT) basis.

DWT analyzes the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. The decomposition of the signal into the different frequency bands is simply obtained by successive high-pass and low-pass filtering of the time domain signal [11], [12].

D. Feature extraction

The number of levels of decomposition is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies required for classification of the signal are retained in the wavelet coefficients [12]. The number of levels was chosen to be 3.

Simple way for extracting feature is to use wavelet coefficients as neural network inputs, directly [14]. But in order to further reduce the dimensionality of the extracted feature vectors, we used statistics over the set of the wavelet coefficients. The following statistical features were used to represent the time–frequency distribution of the ERP signals:

1) Mean of the absolute values of the coefficients.
2) Average power of the wavelet coefficients.
3) Standard deviation of the coefficients.

The typical way for selection the wavelet type is to visually inspect the data first, and if the data are kind of discontinuous, Haar or other sharp wavelet functions are applied; otherwise a smoother wavelet can be employed. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application [12].

In our current work, we compare two wavelet families for feature extraction: Lemarie and Daubechies. Lemarie wavelet shown a good result for extracting the features from schizophrenia EEG signals [13] and also used on the above database for feature extraction from HR and LR matching ERPs [14]. Daubechies wavelet showed a good performance in recognition of alertness level from EEG [12].

E. Artificial Neural Networks and MLP

Artificial neural networks (ANNs) are formed of cells simulating the low-level functions of biological neurons. In ANN, knowledge about the problem is distributed in neurons and connections weights of links between neurons. The neural network has to be trained to adjust the connection weights and biases in order to produce the desired mapping. At the training stage, the feature vectors are applied as input to the network and the network adjusts its variable parameters, the weights and biases, to capture the relationship between the input patterns and outputs. ANNs are particularly useful for complex pattern recognition and classification tasks. The capability of learning from examples, the ability to reproduce arbitrary non-linear functions of input, and the highly parallel and regular structure of ANN make them especially suitable for pattern classification tasks [12], [15].

The most frequently neural networks used are Multi Layer Perceptron (MLP) which is generally supervised-trained with the error back-propagation (BP) algorithm which is used in this work also [16]. One major property of these networks is their ability to find nonlinear surfaces separating the underlying patterns, which is generally considered as an improvement on conventional methods.

F. Network structure and parameter selection

For solving pattern classification problem, MLP employing back-propagation training algorithm was used. Effective training algorithm and better-understood system behaviors are the advantages of this type of neural network. Selection of network input parameters and performance of neural network are important to distinguish between ERPs from HR and LR groups.

A total of 40 individuals from each group's non-matching study were obtained randomly. We used 60 files (30 from HR and 30 from LR) for training and the rest of the 20 files for the testing purpose. The testing data files were never used in the training process.

An ANN (MLP) with a single hidden layer used to classify the signals. We used a simple trial and error approach, changing the number of hidden layers and hidden units to determine the most suitable ANN architecture for different ERP dataset under consideration.

As the conventional BP algorithm with gradient descent and gradient descent with momentum are slow, a few of the modified BP algorithms were tried. Adaptive learning rate BP, resilient BP, Levenberg–Marquardt, and scaled conjugate gradientBP algorithms were examined for training the ANN.

For prevention of over-fitting and reached to best trained ANN, we used Permute Cross Validation (PCV) and early stopping method in training phase [17]. Also we set a procedure that just the ANNs had a good performance on training data in both accuracy and MSE (minimum square error), selected for testing phase. So the ANNs with weak training, rejected. In this procedure after automatically selecting 10 ANNs with best results on training data they go to test phase for testing the test data and then calculated the mean and standard deviation of their accuracy. Accuracy was the correctly classification of test data for each LR and HR groups.

At first we built one ANN for each channels of EEG selected relevant to the [10] that mentioned those channels had good amplitude in Non-matching study. So we have built 22 ANNs using the wavelet statistic features from Non-matching study. After we evaluated each channels independently and got their results, we selected 5 channels that they had best accuracy for classification of two groups. Then we made the final ANNs similar to ANN for each channels and use all of those channels features as input of the network then trained and tested it like the above procedure.

III. RESULTS

Our experiments results shown that channels P8, O2, P5, PO8, and OZ, have better performance in classifications of HR and LR in using of Non-matching study signals, and it was in agreement with the Surface Energy Contour (SEC) map results in [10], shown that the topographic distribution of c247 and the channels with greater amplitude for this component in ERP recording of non-matching study.

Figure 1, illustrated the difference between grand mean of HR and LR groups in Non-matching study for those channels. Also table I show the classification performance for those channels.

In comparing between Lemarie wavelet and Daubechies, we saw that the feature extracted using Lemarie wavelet led to better performance for both HR and LR groups.

The use of variable learning rate backpropagation for MLP network and using just 5 neurons for hidden layer, gave the most successful results in terms of general performance. The output activation is considered to be unknown if all the values at the output node are less than 0.2 .

Also by wavelet statistic features extracted from the above five channels as inputs of just one MLP network with 7 neurons for hidden layer and variable learning rate backpropagation, the results obtained to significant performance for classification between HR and LR groups. Table II provides details of this neural network performance classification of HR and LR groups.

TABLE I
CLASSIFICATION RESULTS ON TESTING DATA SET FOR 5 SELECTED CHANNELS INDIPENDENTLY IN HR AND LR GROUPS

Channels	HR accuracy	LR accuracy
P8	89 ± 3.8	91 ± 3.1
O2	87 ± 3.67	89 ± 3.2
P5	86 ± 3.21	88 ± 3.16
PO8	89 ± 3.8	91 ± 3.4
OZ	90 ± 3.16	92 ± 3.2

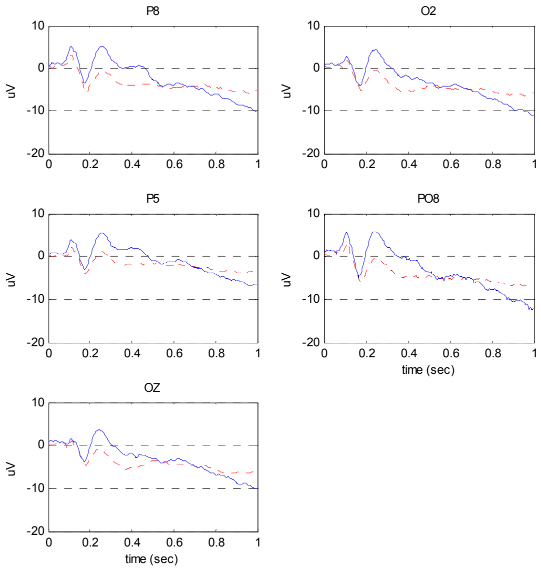

Fig 1. Grand mean ERPs obtained in all subjects for selected channels. Blue solid line shows the LR group and Red dashed line shows the HR group.

IV. Discussion

Our experiment shown that we can use MLP network as a tool for classification of alcoholism ERPs signals for distinguishing between LR and HR groups by choosing the wavelet statistic features instead of total wavelet coefficients that use in other past studies [14], [18]. We reduced the dimensionality of the features without loosing the accuracy, and in agreement with [13] and [14], we saw that by Lemarie wavelet, could reached to better results than Daubechies families.

Also the structure of the ANN reduced, and by 5 hidden layer neurons and Permute Cross Validation and choosing variable learning rate back-propagation for MLP network we shown that we reached to better performance than the above studies. Also if we know the best channels that their involved for distinguishing between two groups in Non-matching study, we can use one ANN instead of ANN for each channels for classification by high performance.

V. Conclusion

All the experiences that we mentioned, shown that there is a linkage between the ERP components and the genetic source for alcoholism. So by study on the ERPs recording in such cases and using computational tools for processing, we can distinguish between subjects at risk for alcoholism at it can help for prevention of abuse of alcohol in sons of alcoholic fathers.

We showed that by choosing better features and using some techniques for better learning of a neural network we can improve the performance of results in classifications individuals at risk for alcoholism. Also we suggest for our future works to use algorithms for choosing the best channels selection and also other processing tools for feature extraction to improve our current works performance.

TABLE II
CLASSIFICATION RESULTS ON TESTING DATA SET FOR FINAL NEURAL NETWORKS WITH FEATURES OF 5 SELECTED CHANNELS AS ITS INPUTS

HR accuracy	LR accuracy
90 ± 3.8	92 ± 3.5

Acknowledgment

The authors wish to thank Lester Ingber for making the contents of the EEG database available on University of California Irvine archive, without restrictions in usage.

References

[1] J.R.M. Copeland, M.T.A. Saleh and D.G. Blazer, *Principles and Practice of Geriatric Psychiatry*, 2nd ed. John Wiley&Sons, 2002, pp.345-348.
[2] "L.J. Bierut, N.L. Saccone, J.P. Rice, A.Goat, " Defining Alcohol-Related Phenotypes in Humans," *Alcohol Research & Health*, vol. 26, no. 3, 2002, pp. 208-213.
[3] B. Porjesz, H. Begleiter, "Event Related Potentials in COA'S", *Alcohol Health & Research Worlds*, vol. 21, no. 3, 1997, pp. 236–240.
[4] P.C. Churchman, R. Moreno, Y. Anez, F. Vergara, "Clinical correlates of quantitative EEG alterations in alcoholic patients", *Clinical Neurophysiology,* no. 117, 2006, pp. 740–751.
[5] B. Porjesz, "The utility of neurophysiological markers in the study of alcoholism", *Clinical Neurophysiolgy,* no. 116, 2005, pp. 993–1018.
[6] C. Reese and J. Polich, "Alcoholism risk and the P300 event-related brain potential: Modality, task, and gender effects", *Brain and Cognition*, no. 53, 2003, pp. 46–57.
[7] K.A. Jones, B. Porjesz, D. Chorlian, M. Rangaswamy, C. Kamarajan, A. Padmanabhapillai, A. Stimus, H. Begleiter, "S-transform time-frequency analysis of P300 reveals deficits in individuals diagnosed with alcoholism", *Clinical Neurophysiology*, no. 117, 2006, pp. 2128–2143.
[8] M. Rangaswamy, B. Porjesz, B.A. Ardekani, S.J. Choi, J.L. Tanabe, K.O. Lim, H. Begleiter. "A functional MRI study of visual oddball: evidence for frontoparietal dysfunction in subjects at risk for alcoholism", *NeuroImage*, vol. 21, 2004, pp. 329-339.
[9] "L. Ingber, (2004, April 15) [Online]. Available at http://kdd.ics.uci.edu/database/eeg/eeg-full.tar"
[10] X.L. Zhang, H. Begleiter, B. Projesz, W. Wang and A. Litke, "Event Related Potentials during Object Recognition Tasks", *Brain Research Bulletin*, Vol. 38, No. 6, 1995, pp. 531-528.
[11] R.Q. Quiroga, H. Garcia, "Single-trial event-related potentials with wavelet denoising", *Clinical Neurophysiology*, no. 114, 2003, 376–390.
[12] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients", *Expert Systems with Applications*, no. 28, 2005, 701–711.
[13] J. Zhu, N. Hazarika, A.C. Tsoi and A.A. Sergejew, "Classification of EEG signals Using wavelet coefficients and an ANN", presented at the pacific workshop on Brain Electric and Magnetic Topographic, Sydney, Australia, 1994.
[14] C.D. Lopes, J.O. Mainardi, M.A. Zaro, and A.A. Susin, "Classification of Event Related Potentials in Indivuiduals at Risk for Alcoholism Using Wavelet Transform an Artificial Neural Network", IEEE 2004 Symposium on Computational Intelligence in Bioinformatics an Computational Biology, La Jolla, 2004, pp. 123-128.
[15] C.E. Vasios, G.K. Matsopoulos, E.M. Ventouras, K.S. Nikita and N. Uzunoglu, "Cross-validation and Neural Network Architecture Selection for the Classification of Intracranial Current Sources", 7th seminar on Neural Network Application in Electrical Engineering, NEUREL-2004.
[16] C. Roberts, J.Gauby and A.Limoge, "Electroencephalogram processing using neural networks", *Clinical Neurophysiology*, 113, pp.695-701, 2002.
[17] N.T. van der Merwe, A.J. Hoffman, "Developing an efficient cross validation strategy to determine classifier performance (CVCP)", International Joint Conference on Neural Networks, Proceedings. IJCNN 2001., Vol. 3 , pp. 1663-1668, 15-19 July 2001
[18] C.D. Lopes, E. Schuler, P.M. Engel, and A.A. Susin, "ERP signal identification of Individuals at Risk for Alcoholism using Learning Vector Quantization Network**",** IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology., 2005.